



HAL
open science

LA LOI ET LES SONDAGES ELECTORAUX

Léo Gerville-Réache

► **To cite this version:**

Léo Gerville-Réache. LA LOI ET LES SONDAGES ELECTORAUX. 49ème Journées de statistique, May 2017, Avignon, France. hal-01531483

HAL Id: hal-01531483

<https://hal.science/hal-01531483>

Submitted on 1 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LA LOI ET LES SONDAGES ELECTORAUX

Léo Gerville-Réache¹

¹Université de Bordeaux, IMS-UMR 5218, France, leo.gerville-reache@u-bordeaux.fr

Résumé. Le 25 avril 2016 ont été adoptées de nouvelles modifications de la Loi n° 77-808 du 19 juillet 1977 relative à la publication et à la diffusion de certains sondages d'opinion. Attendue depuis de longues années, cette évolution nous informe sur les rapports entre le législateur, les instituts de sondage et la statistique. En particulier les termes « échantillon représentatif » et « marges d'erreur » sont aujourd'hui inscrits dans la loi.

Mots-clés. Sondages politiques, Loi de 2016, échantillon représentatif, marges d'erreur, redressement

Abstract. On 25 April 2016 new amendments to Act No. 77-808 of 19 July 1977 concerning the publication of certain opinion polls were adopted. Expected for many years, this evolution informs us about the relations between the legislator, polling institutes and statistics. In particular, terms "representative sample" and "margins of error" are now included in the Act.

Keywords. Opinion polls, representative sample, confidence interval, law.

1 Introduction

A l'heure où ce papier est rédigé, nous sommes à 5 jours du premier tour de l'élection présidentielle 2017. Comme toujours, les sondages, quotidiens, tentent de mesurer l'opinion des électeurs et de quantifier les intentions de vote. Une chose a pourtant changé. Il s'agit de la loi régissant leurs modalités de construction et de publications. En effet, la réflexion sur la réglementation des sondages qui court depuis 2002 et qui a connu une première tentative précise de rédaction et d'adoption en 2010-2011, a trouvé une issue le 25 avril 2016 [2].

On peut lire sur le site de la commission des sondages [1] :

La loi n°2016-508 du 25 avril 2016 apporte plusieurs modifications à la loi du 19 juillet 1977 :

- (1) Elle introduit dans la loi la définition du sondage.*
- (2) Elle étend son champ d'application à tout sondage portant sur le débat électoral.*
- (3) Elle complète la liste des informations devant accompagner la publication des résultats d'un sondage ainsi que celle des informations devant figurer sur la notice.*
- (4) Elle prévoit la mise en ligne des notices sur le site de la commission.*
- (5) Elle précise les pouvoirs de la commission quant à la publication des mises au point qu'elle prononce.*
- (6) Enfin, elle énumère les sanctions applicables.*

La loi modifiée du 25 Avril 2016 comporte un certain nombre d'articles dont les éléments les plus importants (statistiquement parlant) semblent être les suivants [2] :

*Art. 1.- Un sondage est, quelle que soit sa dénomination, **une enquête statistique** visant à donner une indication quantitative, à une date déterminée, des opinions, souhaits, attitudes ou comportements d'une population par l'interrogation d'un échantillon.*

*Les personnes interrogées sont choisies par l'organisme réalisant le sondage de manière à obtenir un **échantillon représentatif** de la population concernée.*

Art. 2.-La première publication ou la première diffusion de tout sondage défini à l'article 1er est

accompagnée des indications suivantes, établies sous la responsabilité de l'organisme qui l'a réalisé :

- Une mention précisant que tout sondage est affecté de marges d'erreur ;
- Les **marges d'erreur** des résultats publiés ou diffusés, le cas échéant par référence à la **méthode aléatoire** ;

Art. 3.-Avant la publication ou la diffusion de tout sondage défini à l'article 1er, l'organisme qui l'a réalisé procède au dépôt auprès de la commission des sondages instituée en application de l'article 5 d'une notice précisant au minimum :

- La méthode selon laquelle les personnes interrogées ont été choisies, le choix et la composition de l'échantillon ;
- La proportion des personnes n'ayant pas répondu à l'ensemble du sondage et à chacune des questions ;
- S'il y a lieu, **les critères de redressement** des résultats bruts du sondage.

S'agit-il d'une révolution ? Cette loi est-elle applicable et appliquée ? Comment les instituts de sondage se sont-ils adaptés ? Voilà des questions sur lesquelles nous souhaitons apporter quelques éléments de réponse.

2 Sur la définition d'un sondage

Le point de l'article 1 qui ne peut nous échapper est celui de l'emploi du terme « échantillon représentatif ».

Les personnes interrogées sont choisies par l'organisme réalisant le sondage de manière à obtenir un échantillon représentatif de la population concernée.

La question de la représentativité de l'échantillon avait fait l'objet de discussions instructives au Sénat en 2010, lors de la première tentative de rédaction et d'adoption d'une loi modificative. Le sénateur Jean-Pierre Sueur, rapporteur de cette loi précisait en séance publique [6] :

Je tiens vraiment à maintenir l'adjectif « représentatif ». Tous les instituts de sondages sont attachés à la notion « d'échantillonnage représentatif ». La communauté mathématique et statistique tient également à cette notion, tout comme la commission des sondages.

Le problème vient du fait que « la communauté mathématique et statistique » n'a nullement trouvé de consensus sur une définition de la représentativité. Par exemple Pascal Ardilly écrivait en 2006 :

« Ce terme ne s'applique pas de façon satisfaisante à un échantillon ».

Mais plus délicat encore, la présente loi ne définit pas le concept. Il est alors bien délicat de se prononcer sur la conformité des sondages avec la loi sur ce point. Les instituts de sondages traitent cette question de manière uniforme dans leurs notices (180 notices en date du 08/02/2016). Par exemple l'IFOP précise [1] :

La représentativité de l'échantillon a été assurée par la méthode des quotas au regard de critères sociodémographiques (sexe, âge de l'individu), socioprofessionnels (profession de l'individu) et géographiques (région administrative, taille d'unité urbaine).

3 Sur les marges d'erreur

La deuxième nouveauté de cette loi porte sur le calcul et la communication des marges d'erreur de chaque résultat du sondage. Si l'on relit le débat de cette loi, il s'agit d'une victoire. En effet, la résistance des sondeurs était forte et la bataille, loin d'être gagnée d'avance. L'argument de base des sondeurs était double :

La méthode des quotas ne permet pas en toute rigueur statistique de calculer des marges d'erreur.

et

La communication de résultats par intervalles nuit à la compréhension du citoyen.

Mais la loi est claire sur ce point. Les sondeurs doivent publier :

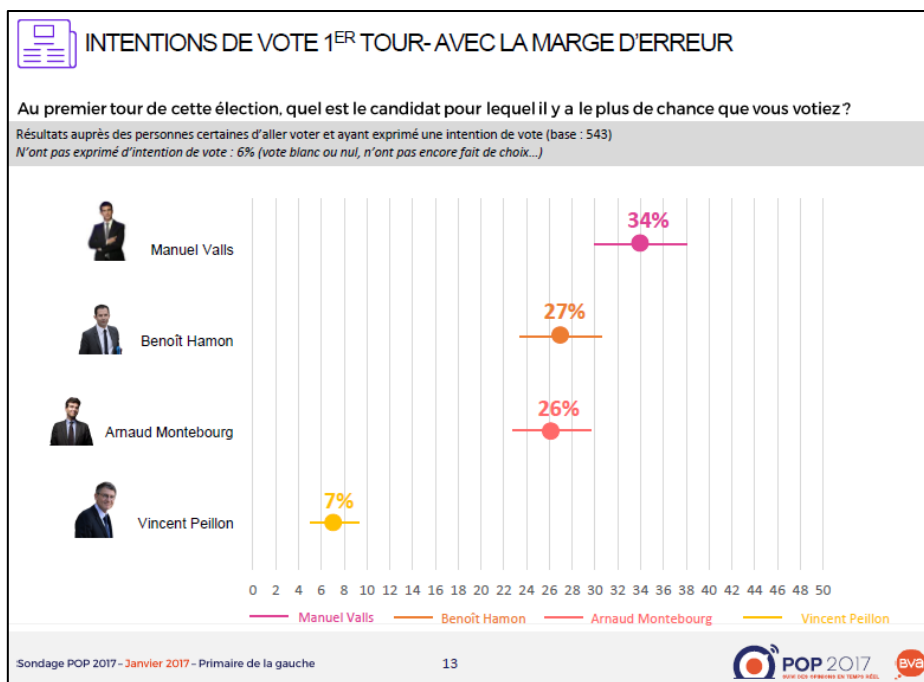
Les marges d'erreur des résultats publiés ou diffusés, le cas échéant par référence à la méthode aléatoire.

Le législateur a donc tranché : peu importe que la méthode de constitution de l'échantillon permette rigoureusement ou pas de calculer des marges d'erreur, les marges d'erreur calculées dans le cadre théorique d'un échantillon simple au hasard seront les références (on peut noter sans y revenir plus avant que cela semble vouloir dire pour le législateur qu'un échantillon simple au hasard est un échantillon représentatif).

L'interprétation des sondeurs de cet article de la loi est également unanime. Les rapports d'enquête, publiés par les sondeurs sont précédés d'un ou plusieurs tableaux généraux sur les marges d'erreurs (des abaques). On peut néanmoins remarquer deux choses. Tout d'abord, ces tableaux étaient déjà présents avant la modification de la loi et le choix du niveau de confiance est systématiquement de 95%. Ce qui est étonnant c'est que la loi ne précise en rien le niveau de confiance qui doit être retenu pour informer sur les marges d'erreur. Les discussions sur la loi n'en faisant pas objet, le consensus sur ce niveau de confiance reste relativement mystérieux.

L'information telle qu'elle est présentée sur les marges d'erreurs est finalement peu pertinente pour le citoyen qui ne voit en réalité que les résultats du sondage et pas les pages d'abaques.

Mais c'était sans compter sur « l'effet primaire de la droite et du centre ». Le décalage entre les sondages et le résultat du premier tour ont conduit deux instituts (BVA et Opinionway) à produire pour les primaires de la gauche enfin un résultat conforme à l'esprit de la loi. L'exemple ci-dessous de BVA n'est pas sans rappeler la proposition que nous faisons il y a 5 ans [4] sur la nécessité d'une charte graphique des résultats des sondages.



Pour autant, cette bonne pratique n'a pas été généralisée et même ces deux instituts n'ont pas poursuivi cet affichage dans les sondages qui ont suivi.

4 Sur les redressements

Le point le plus délicat, pour les sondeurs, est clairement celui qui exige que la notice comporte :

S'il y a lieu, les critères de redressement des résultats bruts du sondage.

En effet, à titre d'extrait, le propos du sénateur Sueur est instructif [6].

La plupart des instituts ont récusé toute publicité des méthodes précises de redressement des sondages, et ce pour deux raisons essentielles :

- d'une part, il convient, selon eux, de ne pas contraindre les instituts de sondage à dévoiler « leurs secrets de fabrication » aux concurrents et, d'une manière générale, au grand public, de même qu'il ne viendrait à l'idée de personne d'imposer à un cuisinier de livrer sa recette précise ;

- d'autre part, si les chiffres bruts étaient publiés en même temps que les chiffres nets, ils sèmeraient le trouble et déclencherait des polémiques sans fin sur l'ampleur des corrections opérées : les instituts pourraient être perçus comme des « bricoleurs », voire des « manipulateurs », à rebours de l'objectif de sérénité et sincérité du débat électoral.

Telle n'est pas la position de vos rapporteurs. Puisque le sondage est revendiqué par les « instituts » comme un produit largement scientifique, il doit être assumé comme tel et être soumis, comme tous les travaux scientifiques, à une obligation de transparence et de démonstration puisque tout travail scientifique doit être vérifiable et reproductible.

Il semble que les instituts aient choisi de répondre à minima à cet article de loi. Les informations sur les redressements sont essentiellement de la forme de celles de l'IFOP [1] :

L'enquête a fait l'objet d'un redressement sociodémographique (données INSEE 2012) à partir des critères ayant servi de quotas (sexe, âge, profession du chef de ménage, niveau d'éducation, région, catégorie d'agglomération) et d'un redressement politique sur la base de la reconstitution du vote suivant : présidentielle 2012 1er tour.

OpinionWay étant le seul institut à fournir des informations plus précises et en particulier les résultats bruts :

Les intentions de vote au premier tour

	PUBLIE	Ensemble des répondants (1664 personnes)				Tout à fait certains d'aller voter (1064 personnes)			
		BRUT	Socio-Demo (SD)	SD & Presid T1 (ST1)	SD & Presid T1 (ST1) & T2 (ST2)	BRUT	Socio-Demo (SD)	SD & Presid T1 (ST1)	SD & Presid T1 (ST1) & T2 (ST2)
Nathalie Arthaud	-	-	-	-	1%	-	-	-	-
Philippe Poutou	1%	1%	1%	1%	1%	1%	1%	1%	1%
Jean-Luc Mélenchon	11%	11%	12%	11%	11%	11%	12%	11%	11%
Yannick Jadot	2%	2%	2%	2%	2%	2%	2%	1%	2%
Benoît Hamon	14%	15%	15%	15%	15%	15%	14%	14%	14%
Emmanuel Macron	23%	25%	25%	25%	25%	23%	23%	24%	23%
François Fillon	20%	22%	20%	20%	20%	21%	20%	20%	20%
Nicolas Dupont-Aignan	3%	3%	3%	2%	2%	3%	3%	3%	3%
Marine Le Pen	26%	21%	22%	24%	23%	24%	25%	26%	26%
N'expriment pas d'intention de vote	18%	14%	18%	23%	22%	9%	10%	12%	12%

*opinionway | Notice à la Commission des sondages – 6 février 2017 | 9

5 Discussion

À l'heure d'un premier bilan sur la loi modificative relative à la publication et à la diffusion de certains sondages d'opinion et son application, le sentiment est partagé.

Sur les exigences de la loi, on peut retenir la volonté soutenue d'éclaircir la production et l'interprétation des sondages. En donnant une définition du terme « sondage » et en l'adossant au terme « d'échantillon représentatif », en augmentant la pression sur la publication des marges d'erreur et les modes de redressement, le législateur a clairement pris la défense du citoyen.

Pour autant, l'absence d'une définition scientifique et/ou légale du concept d'échantillon représentatif reste un problème. Et le problème est d'autant plus délicat qu'en page 27 de ce rapport n°54, est cité un extrait d'un article de Jean-Claude Deville, publié dans la revue *Pour la Science* en juin 2006 [3].

Le statisticien d'origine polonaise Jerzy Neyman publia en 1934 un article fondamental qui montrait qu'un échantillonnage « contrôlé » offrait une justification théorique rigoureuse aux enquêtes par sondage reposant sur cette méthode dite des quotas.

Mais ce n'est pas du tout la citation exacte ! En particulier, le mot « aléatoire » a été omis et la méthode des quotas n'est pas évoquée. Jean-Claude Deville écrivait exactement ceci :

Cela conduisit le statisticien d'origine polonaise Jerzy Neyman à publier en 1934 un article fondamental, qui montrait que seul un échantillonnage aléatoire « contrôlé », en un sens que l'on précisera, offre une justification théorique rigoureuse aux enquêtes par sondage.

Mais pour être très précis, l'extrait qui suit est celui de l'article de Jerzy Neyman de 1934 [5]. L'auteur propose dans sa conclusion une définition statistique du concept de « méthode représentative d'échantillonnage » et compare les sondages aléatoires stratifiés et les sondages par quotas (purposely selected samples).

If there are difficulties in defining the "generally representative sample," I think it is possible to define what should be termed a *representative method of sampling* and a *consistent method of estimation*. These I think may be defined accurately as follows. I should use these words with regard to the method of sampling and to the method of estimation, if they make possible an estimate of the accuracy of the results obtained in the sense of the new form of the problem of estimation, *irrespective of the unknown properties of the population studied*. Thus, if we are interested in a collective character X of a population π and use methods of sampling and of estimation, allowing us to ascribe to every possible sample, Σ , a confidence interval $X_1(\Sigma)$, $X_2(\Sigma)$ such that the frequency of errors in the statements

$$X_1(\Sigma) \leq X \leq X_2(\Sigma) \quad . \quad . \quad . \quad . \quad . \quad (43)$$

does not exceed the limit $1 - \epsilon$ prescribed in advance, *whatever the unknown properties of the population*, I should call the method of sampling representative and the method of estimation consistent. [...]

On the other hand, the consistency of the estimate suggested by Gini and Galvani, based upon a purposely selected sample, depends upon hypotheses which it is impossible to test except by an extensive enquiry.

If these hypotheses are not satisfied, which I think is a rather general case, we are not able to appreciate the accuracy of the results obtained. Thus this is not what I should call a representative method.

Aussi, pour Jerzy Neyman, une méthode d'échantillonnage est représentative si et seulement si on est en mesure d'en déduire scientifiquement les marges d'erreur. Il conclut clairement que la méthode de sélection par quota ne peut aucunement, en toute généralité, revendiquer le terme de représentativité.

D'autre part, le rapport n° 54 contient une annexe 2, rédigée par Pascal Ardilly et intitulé « *Éléments sur la précision des sondages* ».

Il est alors écrit dans le rapport des discussions au sénat page 32 :

Vos rapporteurs estiment qu'il est tout à fait possible de considérer que la méthode des quotas génère des marges d'incertitude similaires à celles obtenues par la méthode aléatoire. Ce point a été confirmé par de nombreuses personnes entendues, en particulier par le statisticien M. Pascal Ardilly dont la contribution, certes complexe, est fort intéressante sur ce point. ».

Mais les propos de Pascal Ardilly ne sont pas cela. Il écrit en particulier dans cette annexe :

L'intention de mesurer et de diffuser l'erreur d'échantillonnage est très louable et doit être encouragée selon nous, mais il faut être clair sur le sens de ces calculs : d'une part ils négligent le biais (sans ce parti pris, il n'y a pas d'intervalle de confiance possible), d'autre part ils résultent bien d'une assimilation à un échantillonnage probabiliste.

En quoi Pascal Ardilly écrit « *qu'il est tout à fait possible de considérer que la méthode des quotas génère des marges d'incertitude similaires à celles obtenues par la méthode aléatoire* » ? Il précise seulement, et à juste titre, que le calcul de marges d'erreur en référence à la méthode aléatoire suppose le parti pris de l'absence de biais. Et à aucun moment dans son écrit, il ne dit souscrire à ce parti pris.

La loi du 25 avril 2016 est une avancée importante dans l'encadrement de la publication et de la diffusion des sondages à caractère politique. Pour autant, malgré les auditions de plusieurs statisticiens, certaines incompréhensions et contre-sens ont conduit le législateur à produire un texte scientifiquement et légalement ambigu.

Néanmoins, parmi les avancées importantes, on peut noter la mise à disposition, en ligne, des notices explicatives de chaque sondage publié. L'institut Opinionway est le seul à présenter un tableau avec les résultats bruts et redressés. Et dans l'émission "Envoyé spécial" du jeudi 13 avril 2017, son directeur général adjoint, Bruno Jeanbart confiait sur le redressement des chiffres bruts :

Chaque sondeur fait ce qu'il veut dans ce domaine, c'est la liberté du sondeur. Puisque le sondeur assume les chiffres qu'il publie, la contrepartie c'est qu'il a une certaine liberté dans la manière dont il va établir les chiffres.

Et dans la même émission, le sénateur Sueur déclarait :

S'il y a des secrets de fabrication par rapport aux redressements, ça veut dire que l'on accepte qu'il y ait une cuisine, pourquoi ne pas le dire, et il y a un certain nombre de cas où on peut le dire, parce que je le sais, où le redressement se fait au doigt mouillé.

La loi du 25 Avril 2016 définit le sondage électoral comme devant être une enquête statistique. A la vue des quelques éléments repris dans ce papier, sur la construction, la rédaction et l'application de cette loi, peut-on scientifiquement considérer que ces sondages électoraux, tels qu'ils sont pratiqués en France, sont des enquêtes statistiques ?

Bibliographie

[1] <http://www.commission-des-sondages.fr>

[2] <https://www.legifrance.gouv.fr/eli/loi/2016/4/25/INTX1527491L/jo/texte>

[3] Deville J. C. (2006). *Peut-on croire aux sondages ?* Pour la science, n° 344, 58-65.

[4] Gerville-Réache L. (2012) *Sondages d'intention de vote : l'estimation des marges d'erreur*, Colloque francophone sur les sondages, Rennes, France.

[5] Neyman J. (1934) *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*, Journal of the Royal Statistical Society, Vol. 97, No. 4, pp. 558-625

[6] Portelli H., Sueur JP. (2010), *Rapport d'information du Sénat n°54 sur les sondages*, Enregistré à la Présidence du Sénat le 20 octobre 2010.